REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1, 1995	September 1, 1992 - September 1, 1995

4. TITLE AND SUBTITLE
Determination of Multiple Sound Sources

5. FUNDING NUMBERS
F49620-92-j-0489
61102F
2313/AS

6. AUTHOR(S)
William A. Yost
Stanley Sheft
Raymond Dye

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Parmly Hearing Institute
Loyola University Chicago
6525 N. Sheridan Rd.
Chicago IL 60626

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR-95
0680

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
Life Sciences
John Tangney
Bolling AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
See Attached

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

See attached

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
U	U	U	U.

Determination of Multiple Sound Sources - **Final Progress Report** - September 1, 1992 to August 31, 1995

AFOSR GRANT: F49620-92-0489

Principal Investigator: William A. Yost, PhD, Parmly Hearing Institute, Loyola University Chicago, Chicago IL, 60626.

Abstract: This three-year research project had the basic aim of understanding the role of binaural hearing in the ability to segregate multiple sound sources in complex sound environments. There were four main projects undertaken over the past three years: 1) To determine the role of binaural cues in sound source identification, 2) To determine the role of spatial separation in processing amplitude modulation, 3) To develop and validate a paradigm for studying analytic and synthetic processing of multiple sound sources, 4) To investigate the role of echoes on the ability of listeners to locate and determine the sources of sound. We found that binaural cues do aid in sound source identification, but that the effects were much greater for three rather than for two sound sources. The ability to process amplitude modulation is aided, but only slightly so, by spatially separating the modulated sources. We developed SALT (Synthetic and Analytic Listening Task) for studying processing of multiple sounds and we validated its use in several binaural and one amplitude modulation experiments. We showed that it is the temporal rather than the spectral properties of a sound and its echo that give rise to the pitch that arises when an echo colors the perception of the original sound source. And we have begun a study of the break down of echo suppression by developing a new set of techniques to study echo suppression as it relates to localization.

Personnel Involved(over all three years):

William A. Yost, PhD--P.I.
Stanley Sheft, PhD--Investigator
Raymond Dye, PhD--Investigator
James Collier--Technician
Richard Beauchamp-Nobbs--Technician

Publications (1992-1995):

1. Yost, William A., Auditory Image Perception and Analysis, Hearing Research 56, 8-19, 1992
2. Yost, William A., Auditory Perception and Sound Source Determination, Current Directions in Psychological Sciences, Vol. 1(6), 1992
3. Yost, W.A., Fay, R.R., Popper, A.(Co-Editors), Psychoacoustics, Springer Verlag, 1993
4. Yost, William A. Overview of Psychoacoustics, in Psychoacoustics,(W.A. Yost, R.R. Fay, and A. Popper, eds), Springer Verlag, 1993
5. Yost, William A and Stanley Sheft. Auditory Perception, in Psychoacoustics, (W.A. Yost, R.R. Fay, and A. Popper, eds), Springer Verlag, 1993
6. Yost, William A. Perceptual Models of Localization, in Proceedings of Perception of Reproduced Sound, Audio Engineering Society, 12, 1993

7. Yost, William A., New Developments in the Study of Spatial Hearing, Audiology Today 6, 9-12, 1994

8. Yost, William A and Sheft, Stanley. Modulation Detection Interference: Across-Spectral Processing and Sound Source Determination, Hearing Research, 79 (1/2), 48-59, 1994

9. Dye, R.H., Yost, W.A., Stellmack, M, and Sheft. A Stimulus Classification Procedure for Assessing the Extent to Which Binaural Processing is Spectrally Analytic or Synthetic, Journal of the Acoustical Society of America 96, 2933-2947, 1994

10. Yost, William A, Dye, R.H., and Sheft, S. Analytic and Synthetic Listening, in Current Topics in Acoustical Research, (J. Menon, Ed.), Research Trends 1, 203-217, 1994.

11. Yost, William A. Tonal Temporal Modulation Transfer Function, Journal of the Acoustical Society of America, in press

12. Yost, William A., Dye, R.H., Sheft, S. The Synthetic-Analytic Task for Modulated Signals, Journal of the Acoustical Society of America, in press

13. Yost, William A. The Cocktail Party Effect: 40 Years Later, in Localization (R. Gilkey and T. Anderson, eds) Erlbaum Press, New Jersey, in press

14. Yost, William A., Dye, R.H., and Sheft, S. A Simulated "Cocktail Party" With Up to Three Sound Sources, Perception and Psychophysics, in press

15. Grange, Tony, and Yost, William A., Effects of off-target interaural differences on dichotic pitch detection, Journal of the Acoustical Society of America, under review

16. Yost, William A., Patterson, R.D., and Sheft, S. A Time Domain Description for the Pitch Strength of Iterated Ripple Noise, Journal of the Acoustical Society of America, in press

17. Sheft, Stanley and Yost, William A., Modulation detection interference with two-tone modulators, Journal of the Acoustical Society of America, under review

18. Dye, Raymond H, Jr., Stellmack, Mark, and Yost, William A., The effect of distractor frequency on judgments of target laterality based on interaural delays, Journal of the Acoustical Society of America, in press

19. Sheft, Stanley and Yost, William A, Role of spatial sues in modulation detection interference, Journal of the Acoustical Society of America, in press

20. Yost, William A., The Pitch of Iterated Rippled Noise, Journal of the Acoustical Society of America, in press

21. Yost, William A., and Dye, Raymond, Binaural Psychophysics, Seminars in Hearing: Binaural Issues in Clinical and Rehabilitative Audiology, Thieme Mdeical Publishers, Inc., in press

22. Sheft, S. (1995). "Effects of carrier asynchrony on modulation detection interference," J. Acoust. Soc. Am. (in press).

23. Stellmack, M. A. and Dye, R. H. (1993). "The Combination of Interaural Information across Frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity" Journal of the Acoustical Society of America 93, 2933-2947.

24. Dye, R. H., Niemiec, A. J., and Stellmack, M. A. (1994). "Lateralization on the Basis of Interaural Envelope Delays: The Effect of Randomizing Component Starting Phase," Journal of the Acoustical Society of America 95.

25. Dye, R. H. (1995). "The Relative Contributions of Targets and Distractors in Judgments of Laterality Based on Interaural Differences of Level," in Binaural and Spatial Hearing (eds. R. H. Gilkey and T. R. Anderson), Lawrence Erlbaum, Associates.

26. Dye, R. H., Stellmack, M. A., and Grange, A. N. (1995). "The Effect of Distractor

Frequency on Judgments of Target Laterality Based on Interaural Delays," Journal of the Acoustical Society of America (in press).

27. Dye, R. H., Stellmack, M. A., and Niemiec, A. J. (1995). "The Discrimination of Interaural Envelope Delays for Waveforms of Comparable Effective Modulation Depth," Hearing Research (in press).

THE APPENDIX - Primary Data Related to the Five Main Projects of the Grant, *Determination of Multiple Sound Sources*, F49620-92-0489

1) Binaural Processing and Sound Source Determination

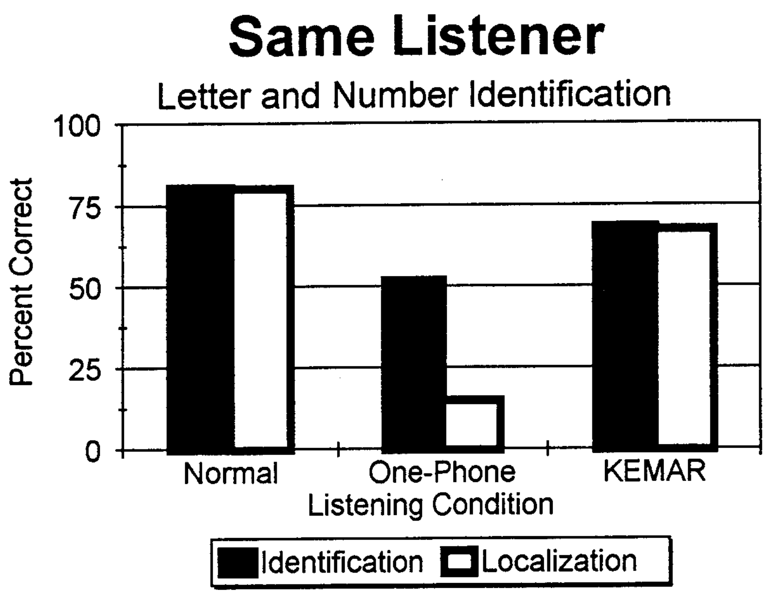


Figure 1. Data from presenting three simultaneous utterances for sound source identification and localization.

Figure 1 above shows the major findings of a study to determine the role of binaural cues in sound source identification. Listeners were presented three simultaneous sets of words and numbers from three of seven possible loudspeakers. They were asked to both identify all the utterances that they could detect (Identification) and then to determine the spatial location of each utterance reported (Localization). They did this in three listening conditions: 1) In the NORMAL listening condition they sat in a special sound-deaden room and used all of the normal cues available when listening to sounds, 2) In the ONE-PHONE condition the sounds were sent by a single microphone placed in the room to a single headphone worn by the listener in a remote soundproof room. This eliminates all binaural cues. 3) In the KEMAR condition the sounds were recorded through the two ears of an acoustic mannikin (KEMAR) and fed to the stereo headphones of the listener in the remote soundproof room. KEMAR maintains many of the binaural cues used by humans to localize sounds. As can be seen, listeners were unable to accurately localize the source of the sound in the ONE-PHONE condition, but were able to identify the sounds in all three conditions. However, there was a substantial drop in identification performance when the sounds were presented over only the one headphone, thus indicating that binaural cues are important for identifying multiple sounds. This is further illustrated by the increase in identification performance in the KEMAR condition in which many of the binaural cues are maintained. The change in performance among these three conditions was much smaller when only two simultaneous sounds were presented, suggesting that binaural cues play a larger role the more complex the listening situation.

2) Amplitude Modulation Processing and Spatial Separation

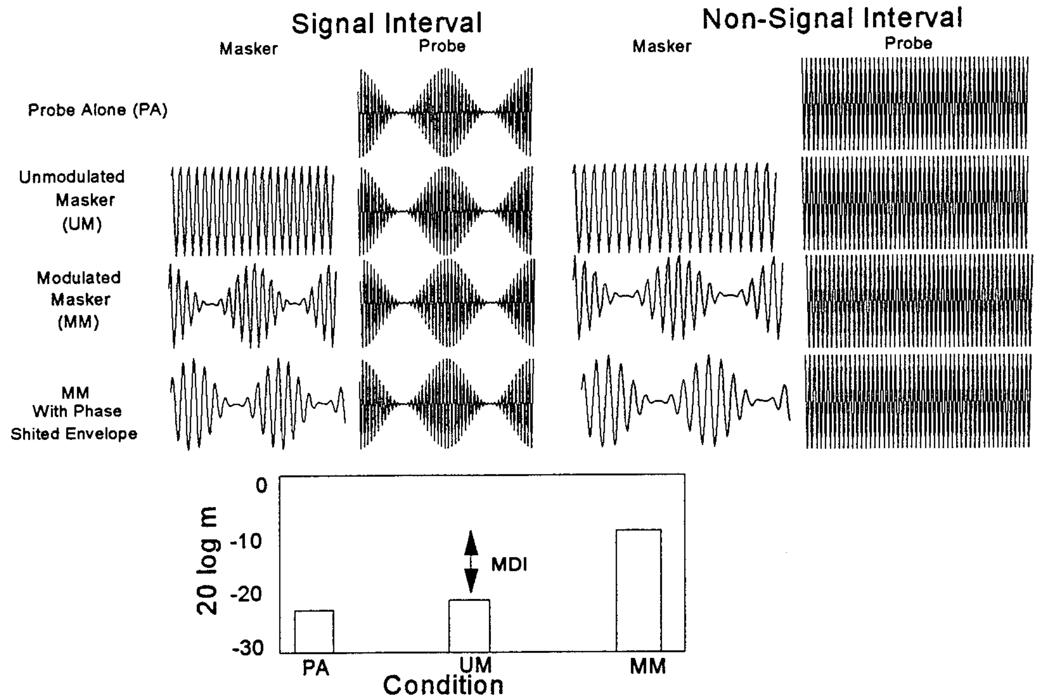
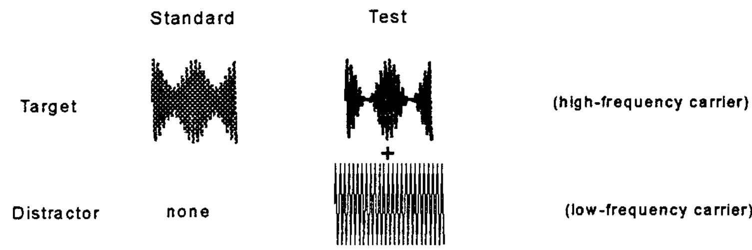


Figure 2. The basic Modulation Detection Interference (MDI) procedure top and result (bottom) when the sounds are all presented to the same loudspeaker or headphone. Listeners are asked to detect the presence of amplitude modulation for the probe occurring in the signal interval in three conditions: Probe Alone (PA) condition in which only the probe is present, Unmodulated Masker (UM) condition in which the unmodulated masker is presented simultaneously with the probe, and Modulator Masker (MM) condition in which the masker is modulated (with its envelope in phase or out of phase with the probe envelope as studied in Experiment I). The typical MDI result is that listeners have low thresholds in the PA condition which change little for the UM condition, but thresholds are elevated drastically in the MM condition. MDI in dB is the difference in thresholds between the MM and UM conditions, where m is the depth of modulation required to detect the modulated probe. MDI is usually expressed in decibels (differences in 20 log m). In this study and in the main condition the probe was presented to one loudspeaker and the masker (modulated or unmodulated) was presented to another loudspeaker. The spatial separation between the loudspeaker was a variable.

When the probe and maskers come from the same location (added at the same loudspeaker) then there is about 11 dB of MDI. That is, the modulation of the probe is difficult to detect when the masker is also modulated. When the probe and masker are spatially separated, the smallest amount of MDI is about 8 dB. So spatially separating the probe and mask can reduce the interference caused by the masker, but the amount of the reduction is small. The explanation for MDI is based on the assumption that the modulated masker and modulated probe form a single auditory source based on their common modulation. As such, it is difficult to process the

modulation of the individual components. Spatially separating the probe and masker provides only a small release from this interference. This suggests that modulation is a more powerful grouping variable in this context than spatial separation. Similar results were obtained when the masker and probes were presented over headphones with differing amounts of interaural differences of time and level.

3) Synthetic and Analytic Listening Task



Possible Responses: "H" because the Target in the Test is greater in modulation depth than it is in the Standard;
"L" because the Distractor in the Test is lower in modulation depth than the Target is in the Standard, and the listener can not attend to only the Target

Thus, an "H" response indicates Analytic Listening and an "L" response Synthetic Listening.

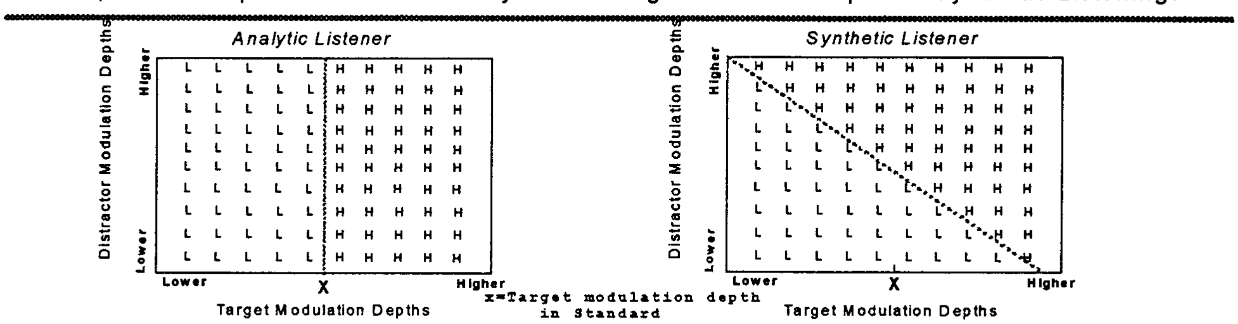


Figure 3. The basic SALT procedure used in a modulation depth detection experiment. On top is a schematic diagram of the stimulus conditions: An amplitude-modulated tonal carrier is presented as the Target during the Standard stimulus and again in the Test stimulus. During the Test, another amplitude-modulated carrier (the Distractor) is simultaneously presented. The listener indicates if the Target stimulus presented during the Test is higher or lower in modulation depth than when it was presented in the Standard. During the Test stimulus, the Target and Distractor are each modulated at one of ten depths of modulation (five higher and five lower than the Target depth of modulation in the Standard). In the example shown at the top, an analytic listener would respond "H" or "higher" since the Target in the Test has a greater depth of modulation than it did during the Standard. A synthetic listener may respond either "L" or "H," since the Target in the Test has a greater depth of modulation and the Distractor a lower depth than that of the Target in the Standard. The lower the depth of modulation of the distractor, the more likely it is that a synthetic listener will respond "L". The response matrices shown below indicate an analytic and synthetic listener for the 100 possible combinations of Target and Distractor modulation depths. The dotted line in each matrix represents the linear boundary partitioning the data into two sets. The slope of this boundary is the ratio of the weights assigned to Targets and Distractors.

The SALT procedure has been used to study modulation processing as indicated in Figure 3 and the lateralization of sounds. In the lateralization task, the target sound of one frequency is presented first with no interaural differences to mark midline or center. The target and distractor are then presented together, where the distractor is a tone of a different frequency. The target and distractor each can take on one of ten values of interaural time (or in some experiments of interaural level), with half of the time differences favoring the right ear and half the left ear. The listener's task is to decide if the target when it is combined with the distractor is left or right of the target when it is presented first alone. Again the data can be partition by the best fitting boundary, and the slope of this boundary can indicate the amount of weight the listener assigns to the target. This procedure has allowed us to show that the auditory system is synthetic in processing sounds that have the same modulation pattern. The procedure also indicates that the auditory system is synthetic in processing interaural differences when sounds differ by more than a critical band in frequency. The procedure as also allowed us to describe in a meaningful manner the individual differences that we measure across listeners.

4) Temporal Basis for the Pitch of a Sound and its Echo

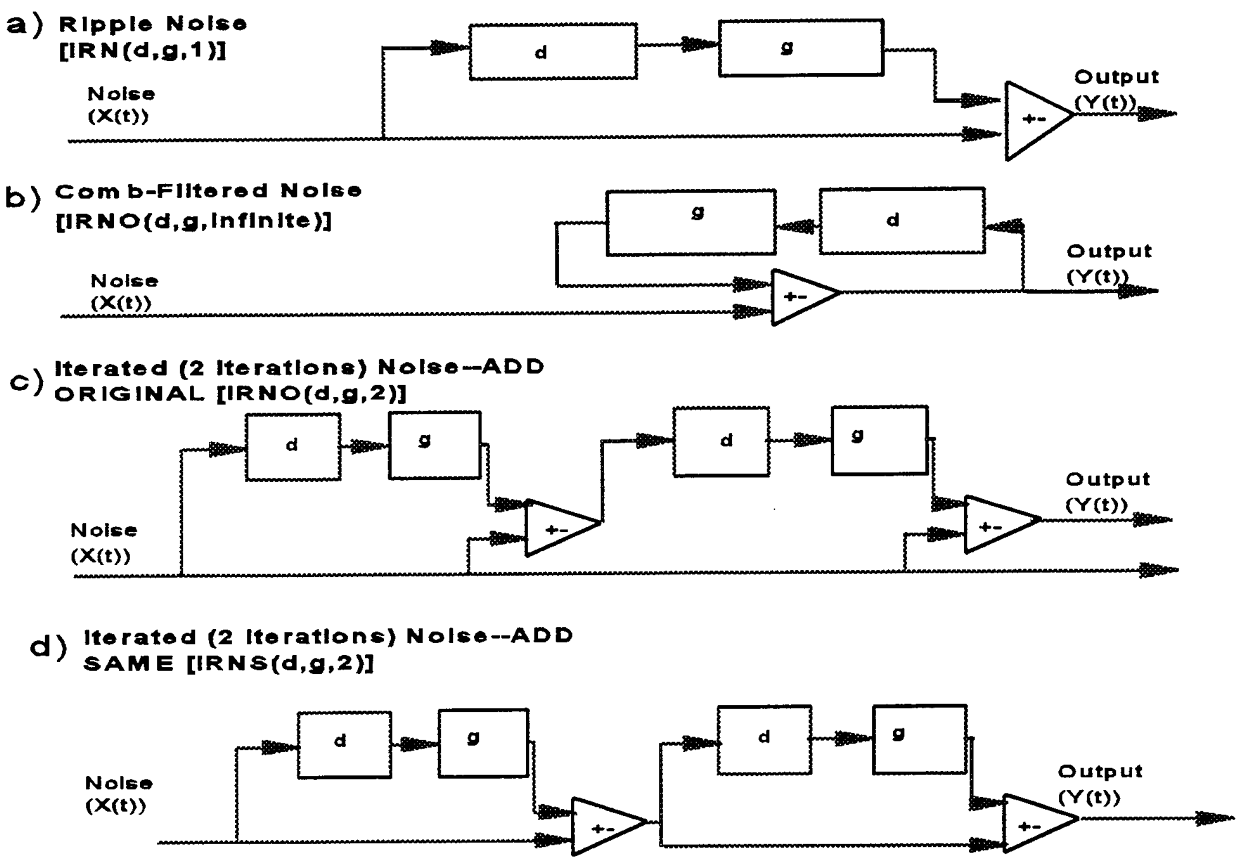


Figure 4 (on the previous page). A set of networks in which a noise input (x(t)) is delayed (d), attenuated (g), and added (+) back to itself one or more times (n iterations). In each case the network simulates the type of situation that might occur when a sound encounters reflective surfaces and the echoes off those surfaces are added to the original sound when it reaches the listener.

The spectra of these sounds have a ripple in amplitude vs. frequency, where the depth of the ripple is determined by the attenuation, g; the spacing between the peaks in the spectrum is determined by the delay, n; and the sharpness of the spectral peaks is determined by the number of stages (iterations) of the delay and add networks. These spectral differences have led many investigators to conclude that the pitch of these stimuli is related to the spectral differences. However, in a series of studies we have shown that the pitch cannot be due to the spectrum, but is much more likely due to the temporal properties of the waveform as shown in Figure 5.

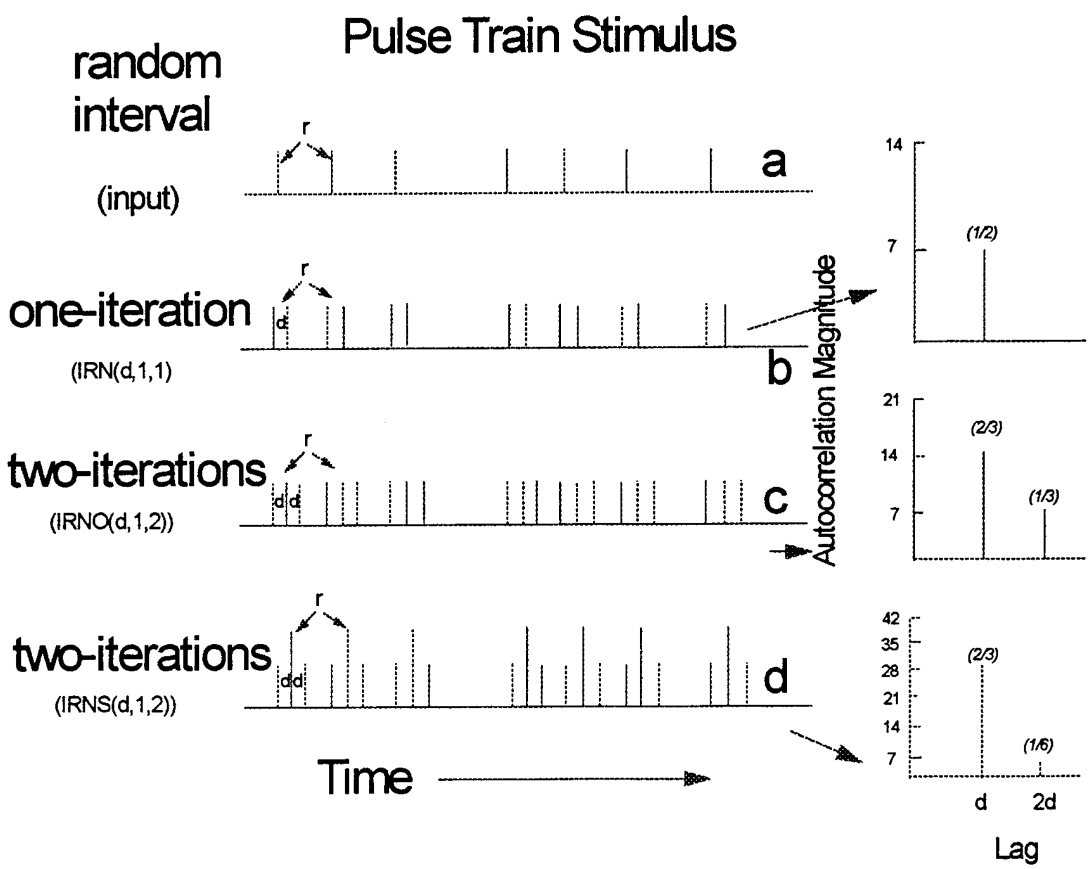


*Figure 5. A noise waveform is represented by a series of random interval pulses. When these pulses are delayed and added back as described in Figure 4 the waveforms contain many intervals of duration d and n*d. This fact can be quantified by calculating the autocorrelation functions for these waveforms as is shown in the right of the figure.*

The peaks in the autocorrelation function (especially the first peak, which indicates the proportion of intervals in the waveform with duration equal to d) can be used to account for the pitch and pitch strength of a noise and its echo. Thus, it is the timing of the fine structure of these sounds that determines the extent to which an echo can influence the perception of the original sound.

5) Break Down of Echo Suppression

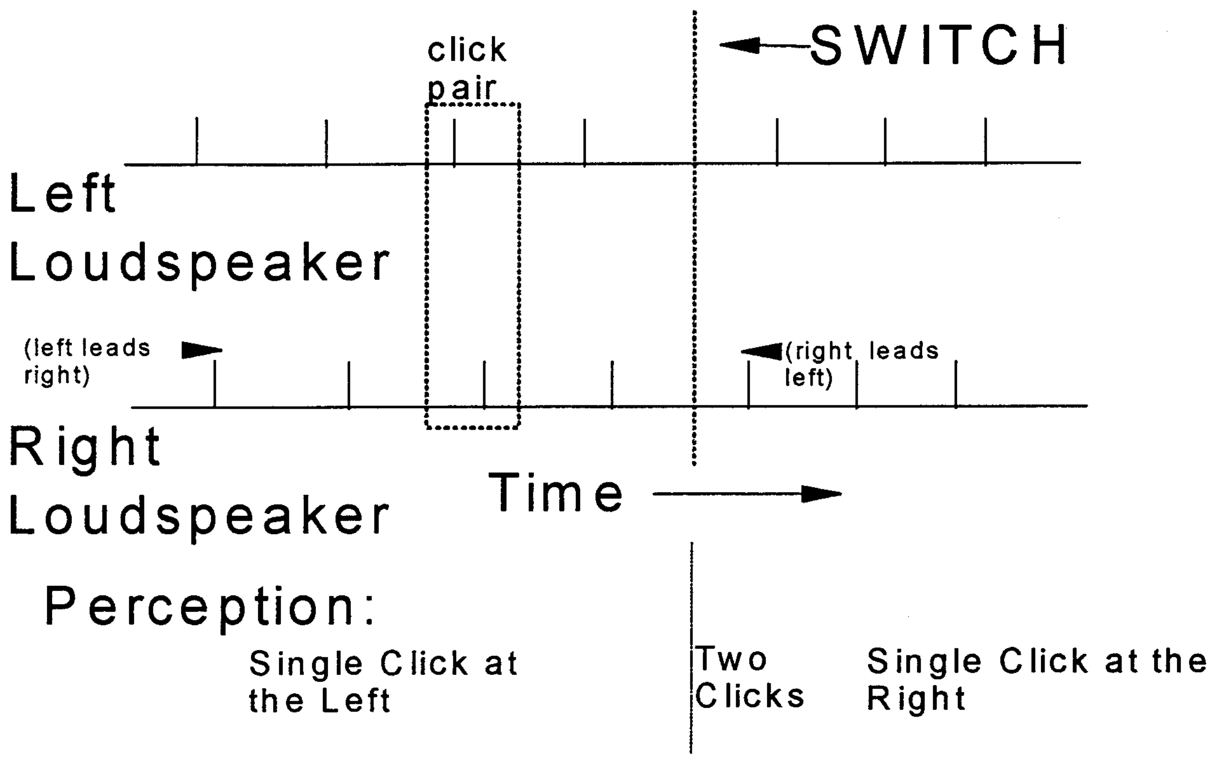


Figure 6. A depiction of the condition in which echo suppression operates and one condition in which there is a break down of the suppression.

In Figure 6 a train of click events or pairs is presented where each click event consists of a click delivered to the left loudspeaker (the source click) followed a few milliseconds by the click being delivered to the right loudspeaker (the echo click). As this click event is presented as a train of events, listeners hear a train of single clicks at the location of the left loudspeaker. This indicates that the trailing or echo click at the right loudspeaker has been suppressed. If after 10 or so click event presentations the order of the source (left) and echo (right) clicks are temporally reversed, listeners perceive clicks coming from both loudspeaker locations for a few moments before echo suppression reestablished itself and the listener again suppresses the trailing or echo click and the listener reports hearing a single click this time at the right loudspeaker which is now the leading or source loudspeaker.

The switch that occurs when the source and echo click are reversed is not a plausible change in the real world. That is, echoes do not suddenly become sources and sources do not suddenly

become echoes. A plausible switch is a situation in which the source and echo both move suddenly. This is what happens when a sound source moves. If this sort of "switch" is introduced, than there is no break down in echo suppression and the listener continues to hear a single click at new location of the lead loudspeaker. We have also shown that these effects operate when there are as many as seven echoes for a single source. In general, any plausible change in the temporal or spatial pattern of sound sources and echoes continues to result in echo suppression, while any implausible switch leads to a temporary break down of echo suppression. Such a break down of echo suppression can occur whenever there is an inappropriate change in the delay of a sound relative to its source. Such circumstances could occur in the use of virtual environments in providing auditory simulations.